

617,3259
7 4 1

[0066] In another embodiment, developers can request transcription of a predetermined number of utterances, e.g., 10,000, from the provider of the zero-footprint development environment (or their affiliates, etc.) for a cost. Then the developer can simply use the accuracy reports without the need for her/him to perform the transcriptions.

[0067] The embodiments described above are illustrative only and not limiting. For example, in other embodiments of the invention, additional steps such as secured login and data encryption may be added to the transcription process. Moreover, data may be displayed in any form that clearly conveys meaningful information during report generation. Other embodiments and modifications to the system and method of the present invention will be apparent to those skilled in the art. Therefore, the present invention is limited only by the appended claims.

1. A method of transcription using a web-based server, the method comprising:

receiving a first request over a network, the first request corresponding to a request to transcribe an utterance;

accessing a set of one or more tuples in response to the first request; and

receiving a second request, the second request corresponding to a human provided transcription of an utterance.

2. The method of claim 1, wherein the first request is generated by a standard web browser.

3. The method of claim 1, wherein the network is the Internet.

4. The method of claim 1, wherein the network is a Virtual Private Network (VPN).

5. The method of claim 1, wherein the network uses an Internet protocol.

6. The method of claim 5, wherein the Internet protocol is Hypertext Transfer Protocol (HTTP).

7. The method of claim 1, wherein each tuple includes: the utterance;

a grammar-in-use during the utterance; and

a recognized result of a speech recognizer of the utterance.

8. The method of claim 7, wherein the tuple is extended to include the human provided transcription of the utterance.

9. The method of claim 1, wherein the set of one or more tuples is aggregated from a larger set of tuples using a first selection criteria.

10. The method of claim 9, wherein aggregation from a larger set of utterance tuples further uses a second selection criteria.

11. The method of claim 9, wherein a first transcriptionist accesses the set of one or more tuples.

12. The method of claim 11, wherein a second transcriptionist accesses a subset of tuples aggregated from the larger set of tuples using the first selection criteria, the set of one or more tuples and the subset of tuples having mutually exclusive tuples.

13. The method of claim 1, wherein the transcription of the utterance includes:

playing an audio definition of the utterance;

defining a text translation of the utterance;

labeling the text translation with audio attributes of the utterance;

labeling the text translation with characterizations of the utterance if present; and

labeling the text translation with utterance anomalies if present.

14. A web-based transcription system, comprising:

a set of one or more stored utterance tuples, each tuple including:

an utterance,

a grammar-in-use during the utterance, and

a recognized result of a speech recognizer from the utterance;

an access system for accessing the set of tuples, the access system including:

a sign-in portion for identifying a transcriptionist and for identifying a subset of the set of tuples,

a persistent label portion for identifying labels consistent across each related portion of the subset of tuples,

a transcription portion for transcribing the utterance associated with each tuple in the subset of tuples; and

an extension system for extending each tuple in the subset of tuples to include the transcribed utterance.

15. The system of claim 14, the access system further including a noise events portion for adding transcription labels to the transcribed utterance defining types of the utterance.

16. The system of claim 14, the access system further including an anomalies portion for adding transcription labels to the transcribed utterance defining qualities of the utterance.

17. The system of claim 14, the access system further including an audio tool for playing the utterance.

18. The system of claim 14, the persistent label portion further including keyboard shortcuts for identifying labels.

19. The system of claim 14, the transcription portion further comprising an auto-complete function for automatically completing a portion of the transcribed utterance.

20. The system of claim 19, the transcription portion further comprising a commonly transcribed utterance list including commonly transcribed utterances beginning with the portion of the transcribed utterance.

21. The system of claim 14, the access system including an information portion for accessing additional information on a portion of the access system.

22. The system of claim 21, wherein the information portion is a help portion and the additional information is help information.

23. A web-based transcription system, comprising:

a set of one or more stored utterance tuples, each tuple including:

an utterance,

a grammar-in-use during the utterance, and

a recognized result of a speech recognizer from the utterance;

means for accessing the set of tuples, including:

- a sign-in portion for identifying a transcriptionist and for identifying a subset of the set of tuples,
- a persistent label portion for identifying labels consistent across each related portion of the subset of tuples,
- a transcription portion for transcribing the utterance associated with each tuple in the subset of tuples; and

means for extending each tuple in the subset of tuples to include the transcribed utterance.

24. The system of claim 23, the transcription portion including a noise events portion for adding transcription labels to the transcribed utterance defining types of the utterance.

25. The system of claim 23, the transcription portion further including an anomalies portion for adding transcription notation to the transcribed utterance defining qualities of the utterance.

26. The system of claim 23, means for accessing further including an audio tool for playing the utterance.

27. The system of claim 23, the persistent label portion further including keyboard shortcuts for identifying labels.

28. The system of claim 23, the transcription portion further comprising an auto-complete function for automatically completing a portion of the transcribed utterance.

29. The system of claim 28, the transcription portion further comprising a commonly transcribed utterance list including commonly transcribed utterances beginning with the portion of the transcribed utterance.

30. The system of claim 23, means for accessing including an information portion for accessing additional information on a portion of the access system.

31. The system of claim 30, wherein the information portion is a help portion and the additional information is help information.

32. A method of drill-down reporting using a web-based system, the method comprising:

- defining a first filter criteria;
- accessing a set of one or more stored utterance tuples meeting the first filter criteria, each tuple including:
 - an utterance,
 - a grammar-in-use during the utterance,
 - a recognized result of a speech recognizer from the utterance, and
 - a transcribed utterance;

providing analysis of the set of tuples in a first standard form of reporting, the first standard form of reporting including internal linking to a first set of support data associated with the set of tuples.

33. The method of claim 32, wherein the set of tuples is aggregated from a larger group of tuples.

34. The method of claim 32, wherein the first filter criteria are defined from user constructed queries.

35. The method of claim 32, the method further comprising tuning of the grammar-in-use in response to the analysis of the set of tuples.

36. The method of claim 32, the method further comprising tuning of a pronunciation of the grammar-in-use in response to the analysis of the set of tuples.

37. A web-based drill-down reporting system, the system comprising:

means for defining a first filter criteria;

means for accessing a set of one or more stored utterance tuples meeting the first filter criteria, each tuple including:

- an utterance,
- a grammar-in-use during the utterance,
- a recognized result of a speech recognizer from the utterance, and
- a transcribed utterance;

means for providing analysis of the set of tuples in a first standard form of reporting, the first standard form of reporting including internal linking to a first set of support data associated with the set of tuples.

38. The system of claim 37, wherein the set of tuples is aggregated from a larger group of tuples.

39. The system of claim 37, wherein the first filter criteria are defined from user constructed queries.

40. The system of claim 37, the method further comprising means for tuning of the grammar-in-use in response to the analysis of the set of tuples.

41. The system of claim 37, the method further comprising means for tuning of a pronunciation of the grammar-in-use in response to the analysis of the set of tuples.

42. A web-based drill-down reporting system, the system comprising:

- a first filter criteria;
- a set of one or more stored utterance tuples meeting the first filter criteria, each tuple including:
 - an utterance,
 - a grammar-in-use during the utterance,
 - a recognized result of a speech recognizer from the utterance, and
 - a transcribed utterance;

means for generating analysis of the set of tuples in a first standard form of reporting, the first standard form of reporting including internal linking to a first set of support data associated with the set of tuples.

43. The system of claim 42, wherein the set of tuples is aggregated from a larger group of tuples.

44. The system of claim 42, wherein the first filter criteria are defined from user constructed queries.

45. The system of claim 42, the method further comprising means for tuning of the grammar-in-use in response to the analysis of the set of tuples.

46. The system of claim 42, the method further comprising means for tuning of a pronunciation of the grammar-in-use in response to the analysis of the set of tuples.

47. A web server system comprising:

- a central processing unit;
- a memory unit; and
- a network interface for sending a message, the message enabling a display screen to display:

Sub. web-based
 Sub. web-based
 Comb. Transcription + Web-based
 Comb. Drill-down + Web-based
 Comb. Sub. web-based

a set of buttons defining audio characteristics, and
an audio tool for playing an audio file.

48. The server system of claim 47, the display screen further enabled to display a submit button for accepting the audio characteristics defined by the set of buttons into a data file.

49. The server system of claim 47, the display screen further enabled to display a text entry box for entering a transcription of the audio file.

50. The server system of claim 49, the display screen further enabled to display a drop-down list of possible text entries for entering into the text entry box.

51. The server system of claim 49, wherein the text entry box is pre-populated with a text entry provided by a speech recognizer.

52. The server system of claim 49, wherein the text entry box is pre-populated with a text entry from a data file associated with the audio file.

53. The server system of claim 47, wherein the set of buttons includes a button defining a gender of a speaker of the audio file.

54. The server system of claim 47, wherein the set of buttons includes a button defining an accent of a speaker of the audio file.

55. The server system of claim 47, wherein the set of buttons includes a button defining a quality of the audio characteristics.

56. The server system of claim 55, wherein the quality is background noise.

57. The server system of claim 55, wherein the quality is noise within a car.

58. The server system of claim 55, wherein the quality is audio information missing at a beginning of the audio file.

59. The server system of claim 55, wherein the quality is audio information missing at an end of the audio file.

60. The server system of claim 55, wherein the quality is side speech.

61. The server system of claim 55, wherein the quality is breath noise.

62. The server system of claim 55, wherein the quality is a sentence fragment.

63. The server system of claim 55, wherein the quality is a touchtone noise.

64. The server system of claim 55, wherein the quality is a hang up noise.

65. The server system of claim 55, wherein the quality is unintelligible speech.

66. The server system of claim 55, wherein the quality is filler speech.

67. The server system of claim 55, wherein the quality is mispronounced speech.

68. The server system of claim 47, the display screen further enabled to display a help tool for providing help for items displayed on the display screen.

69. The server system of claim 68, the help tool providing help for one or more of the set of buttons.

70. The server system of claim 47, the display screen further enabled to display a tutorial tool for providing training information for the server system.

71. A web server system comprising:

- a central processing unit;
- a memory unit; and

a network interface for sending a message, the message enabling a display screen to display:

a grammar, the grammar including an associated link to more information about the grammar, and

an utterance classification associated with the grammar including:

an in-grammar portion defining utterances included in the associated grammar, the in-grammar portion including an associated link to more information about the in-grammar portion, and

an out-of-grammar portion defining utterances outside the associated grammar, the out-of-grammar portion including an associated link to more information about the out-of-grammar portion.

72. The server system of claim 71, wherein the links to more information cause the display screen to display additional information about the associated portions.

73. The server system of claim 70, wherein the additional information is more detailed information about the associated portion.

74. The server system of claim 73, wherein the more detailed information includes associated links to further detailed information about the associated portion.

75. The server system of claim 74, wherein the further detailed information is support data.

76. The server system of claim 74, wherein the further detailed information is one or more audio files.

77. The server system of claim 71, wherein the link to more information about the in-grammar portion causes the display screen to display more detailed information about the in-grammar portion.

78. The server system of claim 77, wherein the more detailed information includes links to further detailed information about the in-grammar portion.

79. The server system of claim 71, the display screen further displaying an in-grammar performance associated with the grammar including:

a correctly accepted portion defining utterances correctly accepted by a speech recognizer, the correctly accepted portion including a link to more information about the correctly accepted portion;

a falsely accepted portion defining utterances incorrectly accepted by the speech recognizer, the falsely accepted portion including a link to more information about the falsely accepted portion; and

a falsely rejected portion defining utterances incorrectly rejected by the speech recognizer, the falsely rejected portion including a link to more information about the falsely rejected portion.

80. The server system of claim 71, the display screen further displaying an out-of-grammar performance associated with the grammar including:

a correctly rejected portion defining utterances correctly rejected by a speech recognizer, the correctly rejected portion including a link to more information about the correctly rejected portion; and

a falsely accepted portion defining utterances incorrectly accepted by the speech recognizer, the falsely accepted

portion including a link to more information about the falsely accepted portion.

81. The server system of claim 71, the display screen further displaying an overall performance associated with the grammar including:

a correctly rejected portion defining utterances correctly rejected by a speech recognizer, the correctly rejected

portion including a link to more information about the correctly rejected portion; and

a falsely accepted portion defining utterances incorrectly accepted by the speech recognizer, the falsely accepted portion including a link to more information about the falsely accepted portion.

* * * * *

5